The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak

THE ROYAL SOCIETY

Received 10 June 2002
Accepted 10 September 2002
Published online 28 November 2002

The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak

D. T. Haydon[1*], M. Chase-Topping[2], D. J. Shaw[2], L. Matthews[2], J. K. Friar[3], J. Wilesmith[4,5] and M. E. J. Woolhouse[2]

[1]*Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada*
[2]*Centre for Tropical Veterinary Medicine, Easter Bush, Roslin, Midlothian EH25 9RG, UK*
[3]*Eoscene Corporation, 700, 5th Avenue, Seattle, Washington 98104, USA*
[4]*Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB, UK*
[5]*Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, University of London, Keppel Street, London WC1E 7HT, UK*

The case-reproduction ratio for the spread of an infectious disease is a critically important concept for understanding dynamics of epidemics and for evaluating impact of control measures on spread of infection. Reliable estimation of this ratio is a problem central to epidemiology and is most often accomplished by fitting dynamic models to data and estimating combinations of parameters that equate to the case-reproduction ratio. Here, we develop a novel parameter-free method that permits direct estimation of the history of transmission events recoverable from detailed observation of a particular epidemic. From these reconstructed 'epidemic trees', case-reproduction ratios can be estimated directly. We develop a bootstrap algorithm that generates percentile intervals for these estimates that shows the procedure to be both precise and robust to possible uncertainties in the historical reconstruction. Identifying and 'pruning' branches from these trees whose occurrence might have been prevented by implementation of more stringent control measures permits estimation of the possible efficacy of these alternative measures. Examination of the cladistic structure of these trees as a function of the distance of each case from its infection source reveals useful insights about the relationship between long-distance transmission events and epidemic size. We demonstrate the utility of these methods by applying them to data from the 2001 foot-and-mouth disease outbreak in the UK.

Keywords: epidemic; contact network; R_0; case-reproduction ratio; foot-and-mouth disease; genealogy

1. INTRODUCTION

The basic reproduction ratio for epidemic processes, R_0, is defined as the average number of secondary cases arising from the introduction of a single primary case into an otherwise fully susceptible population (Kermack & McKendrick 1927; Kendall 1956; Bailey 1957; Bartlett 1960; Anderson & May 1991). Beyond the earliest stages of an epidemic, however, intrinsic factors such as the depletion of susceptibles and extrinsic factors such as the implementation of control measures change the observed case-reproduction ratio to R_t, where R_t is the average number of secondary cases arising from a single case infected at time t (for $t > 0$) and, typically, $R_t < R_0$. Knowledge of R_t is of considerable practical importance when managing an epidemic. If $R_t > 1$ then the epidemic is growing and may be regarded as 'out of control' at time t, indicating that additional control measures may be warranted. However, if $R_t < 1$ then the epidemic is in decline (although this does not necessarily indicate that $R_0 < 1$).

Both R_0 and R_t are most often expressed as combinations of parameters derived from an explicit deterministic model of dynamics of susceptible-infectious-recovered/removed (SIR) cases and estimated by fitting a system of equations to data. However, problems can arise from two sources. First, such an explicitly parametric approach requires that many assumptions be made. Often direct knowledge is required of the total susceptible population size, which in poorly mixed or spatially expanding epidemic situations is difficult to estimate or even define. Commonly the average duration of infection is also required, and many simple SIR formulations make unrealistic assumptions about distributions of infectious periods that can affect the estimates (Anderson & Watson 1980; Lloyd 2001). Second, the formulation is usually deterministic and estimated values of R_0 or R_t are only simple averages (usually over both space and time). Nothing is learned about the actual variance in the distribution from which this average derives—which may be of considerable epidemiological importance (e.g. May & Anderson 1987; Woolhouse *et al.* 1997).

Only very rarely is an epidemic monitored so closely that detailed epidemiological data are collected on almost every infected case and plausible estimates made of probable routes of disease transmission. Estimating the detailed history of a particular epidemic by reconstructing an epidemic trees permits case-reproduction ratios to be estimated from data of this type without having to fit a system of dynamic equations to the data. Assuming that

*Author for correspondence (dhaydon@uoguelph.ca).

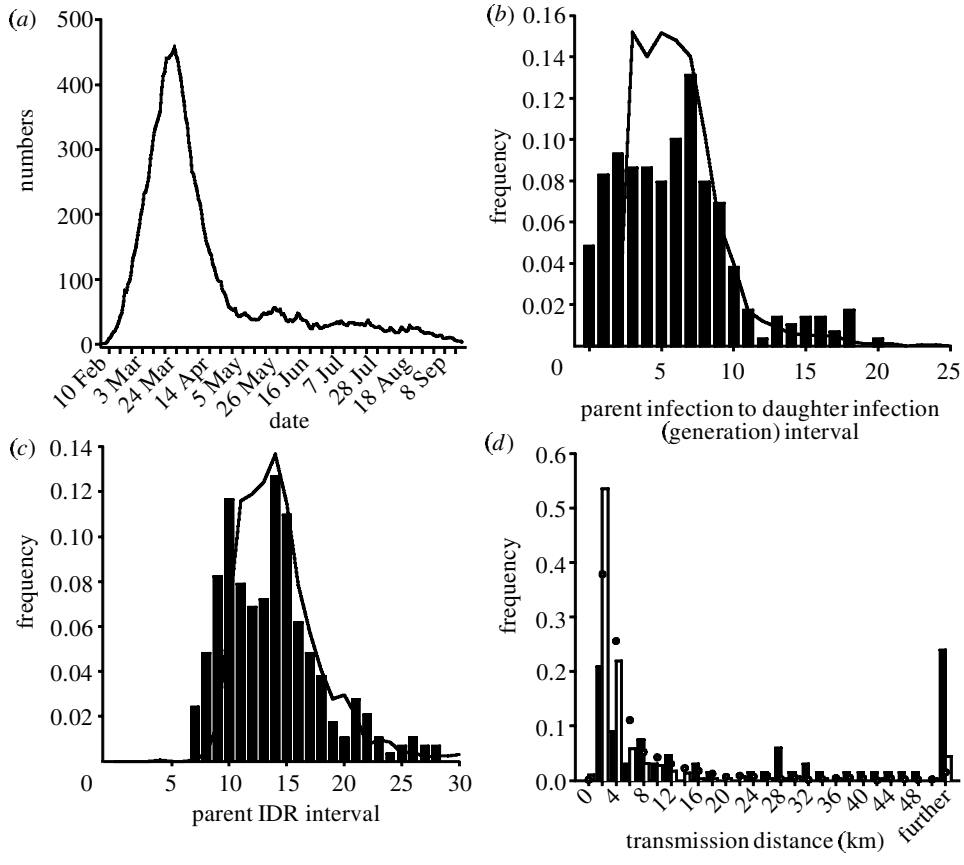

Figure 1. (*a*) The number of properties estimated to be infectious over the course of the epidemic; and (*b*)–(*d*) comparison of data from contact-tracing studies with those inferred with the use of the nearest-neighbour algorithm with $T = 3$ days. (*b*) The distribution of observed generation times (bars) and those with inferred links (line). (*c*) The observed parent IDR intervals (bars) and those with inferred links (line). (*d*) Comparison of the frequencies of infection with distance for links established from contact-tracing pre-NMB (filled bars), from contact-tracing post-NMB (open bars) and inferred links according to the nearest-neighbour reconstruction (open circles).

the resulting parameter estimates are not overly sensitive to exact details of the historical reconstruction, benefits of this approach are twofold. First, estimates of R_t are obtained as directly as possible from the data. Thus, although it is necessary to make some assumptions concerning latency and probability of infection with distance from source, these are minimal. Second, this approach facilitates a detailed description of the spatial and temporal variability manifest in the epidemiological data that is less easily acquired by attempting to fit dynamical models to the underlying process. Furthermore, in contrast to deterministic formulations, an individual-based approach can deal more simply with small sample sizes, permitting easy analysis of both the tail of an outbreak and variation at small spatial scales.

The recent outbreak of foot-and-mouth disease (FMD) virus in the UK provides an example of an outbreak that was monitored sufficiently closely for this sort of approach to be applied (Anon 2001; Morris *et al.* 2001). Cases are infected properties (IPs), the identification of a single infected animal rendering the entire property 'infected'. Familiarity with the pathogenesis of disease permits some estimate of when infection may have first arisen through ageing of lesions. Subsequent culling of all livestock on the property takes place on a known date: therefore each property is infectious for an estimable period of time. The implementation of a ban on livestock movement resulted in mostly local transmission between properties. On a minority of occasions, particularly prior to the introduction of national movement controls, contact-tracing permitted identification of particular properties as the most likely source of infection.

In this paper, we develop a simple methodology that permits reconstruction of 'epidemic-trees' that record which IPs might have or did give rise to which others. These trees permit us to examine three important features of the epidemic: (i) they enable us to estimate R_t and its variance and to observe directly how these quantities vary through space and time; (ii) using the trees, we ask what effect more stringent control measures might have had; and (iii) we examine the influence of long-range transmission events on epidemic size by analysing the cladistic structure of trees with respect to such transmission events. We argue that, in all of these questions, this parameter-free approach can offer a robust alternative to the fitting of deterministic SIR like models to this type of data. We do not dwell on implications of this analysis for control of FMD in the recent UK situation (e.g. Ferguson *et al.* 2001*a*,*b*; Gibbens *et al.* 2001; Keeling *et al.* 2001; Morris *et al.* 2001); such discussion merits a more detailed analysis than space would permit here.

2. METHODS

(a) *The epidemic*

The epidemic, the first case of which was officially confirmed in an abattoir in Essex on 20 Feburary 2001, started on a pig farm in Northumberland that was infected in early February. Retrospective tracing studies indicate that the disease was widespread by the time its presence was confirmed on 20 February. Here, 37 other properties are considered to have been infected by that date (noting that estimates vary from less than 30 to more than 80 properties). A national movement ban (NMB) was imposed on the evening of 23 February, by which time the disease had spread to 40 further properties (again, noting that estimates vary). Between 23 February and 22 September (the date on which the last IP was estimated to have been infected), the epidemic had spread to a further 1948 IPs (making 2026 in total). The epidemic reached properties in most parts of the UK except central and northern Scotland, but the counties of Dumfries and Galloway, Cumbria and Devon were particularly badly affected.

(b) *The data*

Data (last updated on 18 December 2001) on each of 2026 IPs were obtained from the Department for Environment, Food and Rural Affairs (DEFRA) and include: putative dates of infection, reporting (of suspected infection), confirmation, and culling, location and, for 361 IPs, a proposed source of infection identified by contact-tracing (most of which were related to IPs infected early in the epidemic). Thus, sources of infection were identified for 69 of 78 IPs (excluding livestock markets and the putative source) infected prior to the NMB. Knowledge of putative dates of infection and culling permits a simple count to be made of the number of IPs extant in any set of counties on any given day (figure 1*a*).

Contact-tracing is of great importance as it permits estimation of two distributions important to understanding the outbreak dynamics. The first of these distributions is that of generation time, defined to be the interval between the infection time of an IP and the infection time of IPs arising from it (subsequently referred to as 'daughters'). The second distribution is that of the IP infected to daughter reporting intervals (the IDR distribution), which is simply the distribution of time intervals between infection of an IP and reporting of each of its daughters.

(c) *Construction of epidemic trees*

We developed an algorithm that generated a putative source of infection (referred to as a 'parent') for each IP in the following way: when the parent was known from contact-tracing, it was always the assumed infection source. When there was no contact-tracing information available we assumed that the parent itself must have been infected at least T days prior to the infection date of the daughter (Hugh-Jones & Tinline 1976) and extant (not culled) on or before the day of the daughter's infection. Subject to these conditions, the adopted parent was a selected IP from a 'candidate' list, located within a certain distance of the daughter (50 km was chosen as a compromise between an exhaustive candidate list and computational expediency). We adopted three rules for selection of parents from the list of possible candidates: (i) selected parents were simply closest to the daughter (i.e. a single tree was constructed deterministically), we refer to this method as the nearest neighbour algorithm; (ii) they were selected from the candidate list with equal probability; or (iii) parents were selected from the candidate list with probability inversely proportional to their distance (km) from the daughter IP. We fitted an exponential function ($e^{-0.474\,\text{distance}}$) to the distance–infection profiles constructed from parent–daughter links identified from contact-tracing subsequent to the NMB. Markets were considered like any other IP (but only traced transmission events were attributed to them), except for the R_t analysis where daughters of markets were combined with those arising directly from farms and averaged across the farms infected in week zero. With a parent for every IP, it is possible to construct a 'genealogical tree' of the outbreak. Because the second and third methods choose parents randomly from a set of possible candidate parents, each constructed tree will prob-

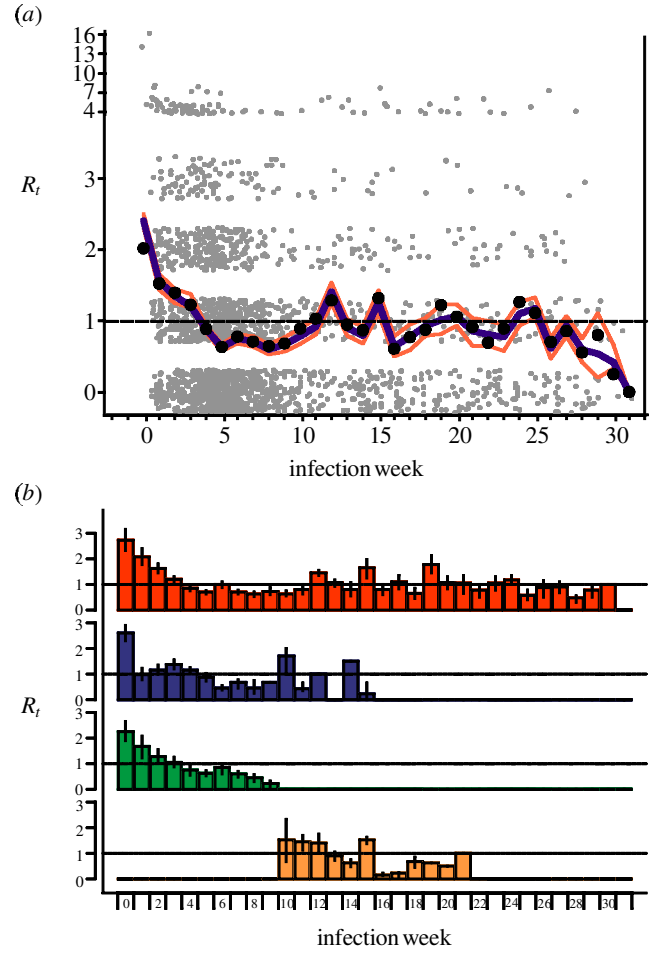

Figure 2. (*a*) R_t values for the whole of the UK from the time of first infection onwards. The black dots show R_t values estimated assuming that parents are always the closest candidate IP. The blue line shows the average R_t values deduced from 500 epidemic trees using method three. Red lines show the bootstrapped 95 PIs. The period between the first infection (7 February 2001) and the imposition of the NMB is the pre-ban period (referred to as 'week zero'), and week 1 starts on 24 February. The small grey circles show the number of daughters for each IP plotted by day of parent infection (random 'jiggle' is added to the *y*-axis to avoid superposition at the integer marks—note the change of scale towards the top of this axis). (*b*) R_t values for selected regions (bars) and PIs obtained from bootstrapping (method three). (Red, Cumbria; blue, Devon; green, Dumfries and Galloway; orange, Settle.) Here, no allowance is made for the effect of changes in culling effort on the reporting of cases; transmission rates are therefore underestimated for weeks three and four (see Woolhouse *et al.* 2001).

ably be different. For each of these latter two methods we constructed 500 epidemic trees and examined various averaged properties of these resampled trees.

(d) *Calculation of R_t values*

Numbers of daughters arising from each parent can be directly counted over the epidemic tree, and R_t estimated by averaging these values within a tree over different time intervals and geographical regions. Here, we present R_t values for each week ($t = 0$–30) of the epidemic up to 22 September over the UK as a whole, and in areas where infection was particularly intense. A distribution of such averages can be obtained by examination of all 500 resampled trees, and 95 percentile intervals (PIs) of this average derived from direct inspection of these distributions.

(e) *Modelling alternative control scenarios*

With a putative epidemic tree containing precise timings and positions of all transmission links, it becomes possible to prune out subsets of links conditional on some assumed control strategy. For example, suppose the NMB had been implemented on 20 February and not 23 February. What impact would this have had on the final epidemic size? This question is addressed by simply 'pruning out' whole branches of the epidemic tree whose existence is conditional on links that arose as a result of transmission events thought to have arisen as a result of animal movements (which we assume to be those that linked daughters to parents who were more than 20 km distant) between 20 and 23 February.

We can also determine the probable effect on final epidemic size had all IPs infected subsequent to imposition of the NMB (as it was implemented on 23 February) been slaughtered within 24 h of reporting. By retrospectively manipulating duration of infectiousness of IPs within a specified tree, certain previously plausible links will become impossible (because daughter infection dates post-date modified culling dates of proposed parents). Elimination of branches within the tree, which are conditional on these now impossible links, permits an estimate of the impact of faster culling times. As we discuss below, because the method as here described assumes no multiple infection pathways, the projected final epidemic sizes are likely to be underestimates.

Results of these pruning processes can be averaged over all 500 tree constructions to obtain minimum estimates (and PIs) of epidemic sizes that might have occurred were these variously different control measures to have been adopted.

(f) *Sub-epidemic analyses*

A large epidemic that spreads predominantly by local transmission will develop around one or more foci of infection, within which the epidemic processes proceed mostly independently. By simply counting up the total number of descendants (over all future epidemic generations) due to an IP that could have arisen purely by local spread (here defined to be that occurring over less than 20 km), it is possible to view the entire epidemic as sets of 'clades' (Hillis *et al.* 1996), within which transmission is solely local in nature. This clade structure can be viewed as a function of the transmission distance at which each IP was itself infected, thereby identifying IPs infected from remote sources that go on to spawn significant sub-epidemics. If the epidemic could have spread entirely locally there will be few of these sub-epidemics, but if longer-range transmission is more frequent, the epidemic should consist of numerous such sub-epidemics. We used nearest-neighbour reconstructions to examine this feature of the epidemic because if these trees are inaccurate they will at least reveal a conservative estimate of the importance of long-distance transmission.

3. RESULTS

Generation time and IDR intervals were calculated for all IPs with known sources of infection that were infected after 23 February ($n = 292$). The distribution of generation times has a mean value of 6.1 days, with standard deviation (s.d.) of 4.6 and skewness 1.9 (figure 1*b*). The IDR distribution has a mean of 13.9 days and s.d. 4.9 (figure 1*c*). Means and variances of both distributions exhibited no significant departure from stationarity subsequent to imposition of the NMB.

Using the nearest-neighbour algorithm and setting $T = 3$ days, we obtained the best-fitting average generation time for inferred links (from daughters infected between 24 February and 22 September) of 6.3 days. Observed and 'simulated' distributions of generation times and IDR intervals matched well (figure 1*b*,*c*). The distribution of infection link distances between parents and daughters generated by the nearest-neighbour algorithm also matched the observed distribution of infection link distances closely (figure 1*d*). The very local nature of the spread of infection is indicated by the distance to the nearest potential (or actually identified) parent subsequent to the NMB, the median of which was just 2.5 km (and mean 7.5 km). By contrast, prior to the NMB this median distance was 14.9 km (and mean 65.3 km).

Nationwide R_t values for the epidemic are shown in figure 2*a*. PIs from methods two and three were very similar and we only report those from method three. Averages from bootstrapped trees and results from deterministically assembled nearest-neighbour reconstructions are very similar post-NMB. It is clear that, nationally, the disease was out of control ($R_t > 1$) until the fourth week following the NMB, but came under control ($R_t < 1$) by week four or five (as date of infection; transmissions occurred one generation (i.e. 6–7 days) later), falling R_t values coinciding with the implementation of shorter reporting to slaughter times and more intensive pre-emptive culling. The subsequent fluctuation in R_t values on the national scale reflects the emergence of uncontrolled smaller outbreaks in Devon, Yorkshire and to a lesser extent Cumbria. Figure 2*b* shows estimates for sub-regions of the epidemic, indicating how infection activity changed in different parts of the country at different times. PIs remain narrow, even with the smaller number of parents in these regions.

The distribution of numbers of daughters as deduced from the nearest-neighbour algorithm for the pre-NMB period are highly over-dispersed relative to a Poisson distribution, having a variance-to-mean ratio of 3.98. After the NMB, the variance-to-mean ratio falls to 1.52 (figure 3).

Removing all infections occurring over a distance of 20 km between 20 and 23 February inclusive (simulating the imposition of an earlier NMB) results in the reduction in epidemic size to an average of 793 cases (figures 4*a*; 95 PI: 654–1012). During this 4-day period, 17 IPs were infected from sources more than 20 km distant and 16 of these IPs have traced sources of infection, 13 of which were markets. Had the NMB been imposed one day later

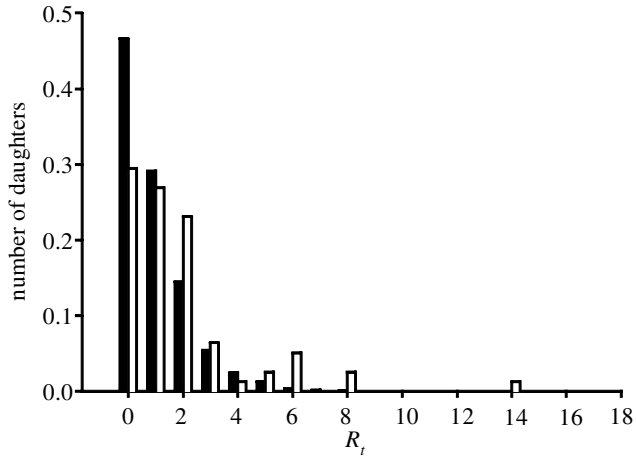

Figure 3. The distribution of numbers of daughters as deduced for all IPs up to 22 September for the pre-NMB (open bars) and post-NMB (filled bars) periods as inferred from the nearest-neighbour tree reconstruction.

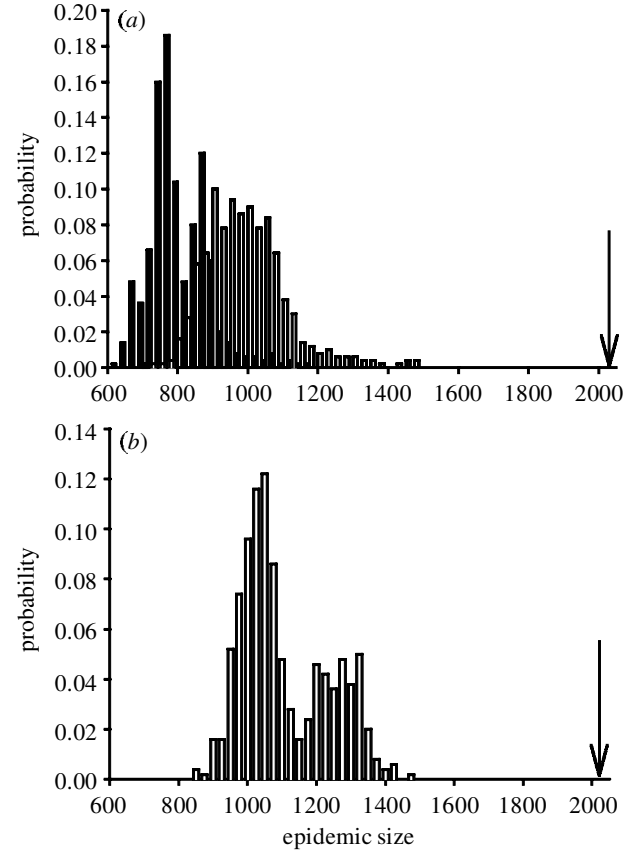

Figure 4. The distribution of number of properties to have been infected during the epidemic predicted from pruning epidemic trees subject to alternate control strategies. (*a*) Imposition of NMB on 20 February (black bars) and 21 February (grey bars) (assuming this to have eliminated all transmission events over 20 km distant). (*b*) Assuming that all IPs reported subsequent to 23 February were culled by the end of the day following their reporting. Arrows indicate the actual final size of the epidemic.

on 21 February, we estimate the expected epidemic size to have been 977 cases (figure 4*a*; 95 PI: 794–1285). Observed reporting-to-slaughter times subsequent to the NMB averaged 1.23 days, with 74% of IPs culled by the day subsequent to reporting. However, if culling dates of IPs infected subsequent to the NMB are all back-dated to the IP's reporting date plus one day, final epidemic size is reduced to an average of 1093 (figure 4*b*; 95 PI: 911–1340). For reasons discussed below, these predicted epidemic sizes are likely to be the smallest that could arise from these alternative control scenarios.

Prior to imposition of the NMB, nearest-neighbour reconstructions identified nine IPs infected from a source more than 20 km away that went on to infect (locally, and either directly or indirectly) between 10 and 50 further properties, and 11 that went on to leave over 50 locally infected descendants. These transmissions were responsible for the early and widespread dissemination of infection that laid the foundations of the epidemic. Subsequent to the NMB, these figures became 18 (infecting on average 18 IPs each) and two, respectively (figure 5*a*), indicating that the outbreak made only modest subsequent incursions into areas uninfected prior to the NMB. After the NMB, the reconstruction suggests that 173 more IPs were infected at distances exceeding 20 km, but each of these IPs went on to leave less than 10 locally infected descendants. In this way, parent IPs may be viewed as the roots of clades representing sub-epidemics within the tree. According to the nearest-neighbour tree, over 80% (1675) of IPs infected belong to 28 clades composed of 10 or more IPs (10 of which were seeded prior to 20 February and 18 of which were seeded prior to imposition of the NMB). The temporal 'evolution' of these clades may be viewed over time showing how each contributed to the overall epidemic size at different times (figure 5*b*). By 22 September, a nearest-neighbour tree contained 38 generations of IPs, with the largest being the 10th generation containing 199 IPs.

4. DISCUSSION

We have described a straightforward yet robust way of reconstructing an epidemic history. This provides a very natural way of viewing the dynamics of a particular epidemic and permits estimation of case-reproduction ratios very directly from the data with a minimum number of assumptions. A method for generating PIs indicates that these estimates are also robust to uncertainties in the historical reconstruction. Accurate knowledge of these parameters is vital if adopted control measures are to be objectively appraised. The bootstrapping procedure described also suggests a method for retrospectively estimating the impact of more stringent control scenarios through the 'pruning' of certain tree branches, whose existence is conditional on transmission events that might have been prevented under these alternative control scenarios. We stress that this method does not attempt to estimate the process underlying the epidemic (cf. Stegeman *et al.* 1999); it is true that the same process, through chance alone, can give rise to a range of actual epidemics (Keeling *et al.* 2001). However, the realized epidemic remains our best source of parameter estimates and so the R_t values estimated here should be consistent with those estimated by fitting dynamical models, and this is indeed the case (cf. Ferguson *et al.* 2001*a*). The advantage of this method is that it makes fewer assumptions, and is there-

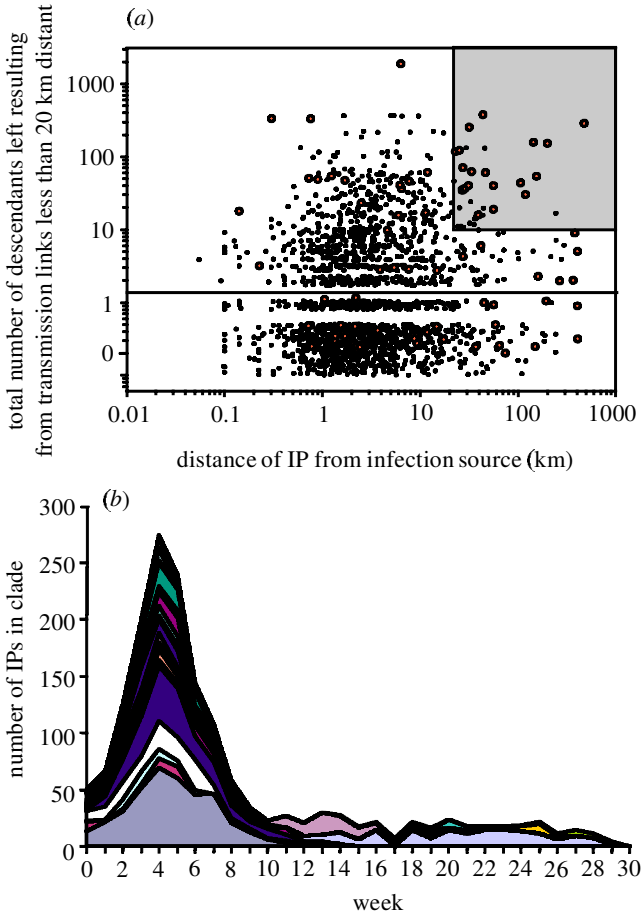

Figure 5. (*a*) A representation of the clade structure of the epidemic tree as a function of long-distance transmission. The graph shows all of the epidemiological descendants from each IP, arising over parent–daughter transmission distances less than 20 km plotted against the distance of the IP from its own (minimum possible) source of infection. (Black dots, IPs infected post-NMB; red dots, IPs infected pre-NMB.) Thus IPs in the grey box in the top-right portion of the graph were infected by sources greater than 20 km away and went on to leave more than 10 direct and indirect descendants, each of which arose through local transmission events less than 20 km distant. Some random jiggle factor has been added to the *y*-axis values to avoid superimposition of points. (*b*) The contribution of each of the 28 clades (each indicated by a different colour) composed of at least 10 IPs within the nearest-neighbour tree as they arise over the course of the epidemic (for example, the purple clade arising predominantly between weeks 12–18 arose in North Yorkshire, and was initiated by 26 April 2001).

fore more robust. One practical lesson learnt from modelling the UK FMD outbreak was the value of using different analytical approaches in parallel.

This method of estimating R_t is robust if it is possible to relate sets of daughter outbreaks to sets of parents. This approach benefits statistically from the high degree of disease surveillance as it is likely that the set of IPs has been exhaustively sampled, and that R_t has therefore been examined using data from all IPs, and not just a sample of them. While numbers of daughters will clearly be sensitive to uncertainties in exactly which daughters are linked to which parents, the mean value for any time-period averages over these uncertainties, thereby providing relatively robust average estimates. Indeed, if the algorithm is modified so that even known links are instead estimated, the analysis changes very little (results not shown).

The tree-pruning algorithms do make important assumptions, the strongest of which is that properties are not subject to multiple infection. If an IP is infected from more than one source it is less likely to be pruned out of the tree as a result of removal of links that could not have occurred under alternative control scenarios. It is therefore necessary to view final epidemic sizes predicted from these alternative control scenarios as minimum estimates conditional on the implementation of the neighbourhood culling that took place. The impact of multiple infection on predictions from bootstrapped tree-pruning can of course be investigated by including multiple infection paths in the reconstructed trees. However, the pruning performed in the analyses presented here remove infections from mostly early in the epidemic (prior to imposition of the NMB and when reporting to culling times were longest), which is also when multiply infected IPs are likely to have been less common due to the more dispersed nature of infections in the early stages of the epidemic. Biases and inaccuracies in the estimated dates of infection could also affect conclusions using this approach: in particular, the predicted impact of earlier imposition of a NMB will be sensitive to exactly which properties were infected by animal movements between 20 and 23 February.

This form of analysis permits epidemiological dynamics to be interpreted directly in terms of standard demographic theory because the distribution of generation time is the product of the life-history parameters, l_x (here interpreted as the culling rate as a function of time since infection, x) and m_x (the infection rate as a function of time since infection, x) corresponding respectively to conventional survivorship and birth rates at age x (Woolhouse & Anderson 1997). Cohort generation time (here estimated at 6.3 days) is estimated by $\kappa_1 = \Sigma_{x=0}^{\infty} x m_x l_x$ and $R_t = \Sigma_{x=0}^{\infty} m_x l_x$. The intrinsic rate of increase (or decrease) of the epidemic process at a stable age distribution is given by $r_t \cong \kappa_1^{-1}[\ln R_t + r_t^2\kappa_2/2 - r_t^3\kappa_3/3]$ (MacArthur & Wilson 1967), where κ_2 is the variance and κ_3 the skewness of the generation time distribution. Higher moments of this distribution will not contribute much to the intrinsic rate of increase, as long as r is quite small (for example, with $R_t = 2$, and $\kappa_1 = 6.1$, eliminating both the variance and skewness completely decreases r only from 0.127 to 0.114), suggesting that all control efforts should be directed at reducing R_t.

Management of the epidemic will be facilitated if the time interval between occurrence and recognition of changes in the epidemiological dynamics is reduced to a minimum (Woolhouse *et al.* 2001). Because it may take several weeks for all 'daughters' of any parent to be reported (the IDR distribution indicates that only 60% of infections due to any IP were reported within two weeks of its own infection, rising to over 93% by three weeks), any attempts to measure the number of daughters arising from more recently infected IPs may underestimate this number by virtue of failing to account for 'daughters yet to appear'. One way to overcome this problem is to compensate estimates of numbers of daughters from more recent IPs with the use of the IDR distribution, which

indicates the expected proportion of daughters yet to be reported for any given time since parent infection.

Our analysis highlights various important but characteristic features of FMD virus epidemiology. There is considerable variation in the case-reproduction ratio of each IP, which arises in this analysis as a result of heterogeneity in the spatial distribution of properties (Hugh-Jones 1972), and it is likely that this variation has been underestimated here as we have ignored probable differences in susceptibility and infectiousness of IPs arising from different stock (e.g. Donaldson *et al.* 2001). While long-distance infections apparently occurred subsequent to the NMB, they appear rare relative to local spread. Analysis of trees constructed assuming that parents were the nearest infectious property to daughters reveals little need to invoke long-range transmission after the imposition of the NMB, and that only a small fraction of IPs may have arisen as a consequence of them. That 54% (20 out of 37) of long-distance transmission (more than 20 km) went on to produce sub-epidemics containing 10 or more IPs prior to the NMB, compared with only 10% (20 out of 193) after the NMB, illustrates the great variability observed in case reproduction ratios and the low establishment rate of FMD virus in the face of the adopted countermeasures (HMSO 1954; Tinline 1972). This conclusion is underscored by results of the pruning simulations that strongly suggest that movement bans must be imposed at the earliest possible opportunity and reiterates the necessity of ensuring that resources do not limit the rapidity with which infected herds can be culled (e.g. Haydon *et al.* 1997; Howard & Donnelly 2000; Morris *et al.* 2001).

The authors are grateful to the Wellcome Trust and DEFRA for financial support, and to the Epidemiology Team at the VLA, particularly Jane Gibbens and Kate Sharp. We thank Alun Lloyd, Loeske Kruuk and Scott McEwen for valuable comments on the manuscript, and three anonymous reviewers for particularly helpful reviews.

REFERENCES

Anderson, D. & Watson, R. 1980 On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika* **67**, 191–198.

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control*. Oxford University Press.

Anon 2001 Foot-and-mouth disease: two weeks and counting. *Vet. Rec.* **148**, 286–288.

Bailey, N. T. 1957 *The mathematical theory of epidemics*. London: Griffin.

Bartlett, M. S. 1960 *Stochastic population models in ecology and epidemiology*. London: Menthuen.

Donaldson, A. I., Alexandersen, S., Sorensen, H. & Mikkelsen, T. 2001 Relative risks of the uncontrollable (airborne) spread of FMD by different species. *Vet. Rec.* **148**, 602–604.

Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. 2001*a* The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155–1160.

Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. 2001*b* Transmission intensity and impact of control policies on the foot-and-mouth epidemic in Great Britain. *Nature* **413**, 542–548.

Gibbens, J. C., Sharpe, C. E., Wilesmith, J. W., Mansley, L. M., Michalopoulou, E., Ryan, J. B. M. & Hudson, M. 2001 Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *Vet. Rec.* **149**, 729–743.

Haydon, D. T., Woolhouse, M. E. J. & Kitching, R. P. 1997 An analysis of foot-and-mouth disease epidemics in the UK. *IMA J. Math. Applicat. Med. Biol.* **14**, 1–9.

Hillis, D. M., Moritz, C. & Mable, B. K. 1996 *Molecular systematics*. Sunderland, MA: Sinauer.

HMSO 1954 *Report of the departmental committee on foot-and-mouth disease 1952–1954*. London: HMSO.

Howard, S. C. & Donnelly, C. A. 2000 The importance of immediate destruction in epidemics of foot-and-mouth disease. *Res. Vet. Sci.* **69**, 189–196.

Hugh-Jones, M. E. 1972 Epidemiological studies on the 1967–1968 foot-and-mouth epidemic: attack rates and cattle density. *Res. Vet. Sci.* **13**, 411–417.

Hugh-Jones, M. E. & Tinline, R. 1976 Studies on the 1967–1968 foot-and-mouth disease epidemic: incubation period and herd serial interval. *J. Hygiene* **77**, 141–153.

Keeling, M. J., Woolhouse, M. E. J., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J. & Grenfell, B. T. 2001 Dynamics of the 2001 UK foot-and-mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817.

Kendall, D. G. 1956 Deterministic and stochastic epidemics on closed populations. In *Proc. 3rd Berkeley Symp. Math. Statistics and Probability*, vol. 4, pp. 149–165. Berkeley: University of California Press.

Kermack, W. O. & McKendrick, A. G. 1927 A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond.* A **115**, 700–721.

Lloyd, A. L. 2001 The dependence of viral parameter estimates on the assumed viral life cycle: limitations of studies of viral load data. *Proc. R. Soc. Lond.* B **268**, 847–854. (DOI 10.1098/rspb.2000.1572.)

MacArthur, R. H. & Wilson, E. O. 1967 *The theory of island biogeography. Monographs in population biology*. Princeton University Press.

May, R. M. & Anderson, R. M. 1987 Transmission dynamics of HIV infection. *Nature* **326**, 137–142.

Morris, R. S., Wilesmith, J. W., Stern, M. W., Sanson, R. L. & Stevenson, M. A. 2001 Predictive spatial modeling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain, 2001. *Vet. Rec.* **149**, 137–144.

Stegeman, A., Elbers, A. R. W., Smak, J. & De Jong, M. C. M. 1999 Quantification of the transmission of classical swine fever virus between herds during the 1997–1998 epidemic in the Netherlands. *Prev. Vet. Med.* **42**, 219–234.

Tinline, R. 1972 A simulation study of the 1967–1968 foot-and-mouth epizootic in Great Britain. PhD thesis, University of Bristol, UK.

Woolhouse, M. E. J. & Anderson, R. M. 1997 Understanding the epidemiology of BSE. *Trends Microbiol.* **5**, 421–424.

Woolhouse, M. E. J. (and 12 others) 1997 Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342.

Woolhouse, M. E. J. (and 11 others) 2001 Epidemiology: foot-and-mouth disease under control in the UK. *Nature* **411**, 258–259.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.